

Additional codes and people in the *British Business Census of Entrepreneurs* (BBCE) not available through I-CeM

Carry van Lieshout, Robert J. Bennett and Harry Smith

cv313@cam.ac.uk rjb7@cam.ac.uk hjs57@cam.ac.uk

Working Paper 21:
Working paper series from ESRC project ES/M010953:
Drivers of Entrepreneurship and Small Businesses

University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure, Downing Place, Cambridge, CB2 3EN, UK.

October 2019

Comments are welcomed on this paper: contact the authors as above.

© Carry van Lieshout, Robert J. Bennett, and Harry Smith, University of Cambridge, members of the Cambridge Group for the History of Population and Social Structure assert their legal and moral rights to be identified as the authors of this paper; it may be referenced provided full acknowledgement is made: *Cite* (Harvard format):

van Lieshout, Carry, Bennett, Robert, J., and Smith, Harry (2019) *Additional codes and people in the British Business Census of Entrepreneurs (BBCE) not available through I-CeM*. Working Paper 21: ESRC project ES/M010953: ‘Drivers of Entrepreneurship and Small Businesses’, University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure.

Keywords: Entrepreneurship, Employers, Self-employment, Census, BBCE, I-CeM

JEL Codes: L26, L25, D13, D22

Additional codes and people in the *British Business Census of Entrepreneurs* (BBCE) not available through I-CeM

Carry van Lieshout, Robert J. Bennett and Harry Smith

Working Paper 21: ESRC project ES/M010953: Drivers of Entrepreneurship and Small Businesses, University of Cambridge

1. Introduction

This paper provides information on data corrections and additions to the digital version of the census data for England and Wales and Scotland that is held within the UK Data Archive (UKDA) data deposit: *The Integrated Census Microdata* (I-CeM) for 1851-1911 (Higgs et al., 2014, 2015; Schürer et al., 2016). These data corrections have been identified during the preparation and subsequent data deposit of the *British Business Census of Entrepreneurs* (BBCE) at UKDA. For BBCE, the transcripts of the censuses, and coding of individuals, are mostly derived from I-CeM, but BBCE adds to I-CeM the identification and coding of entrepreneurs, plus data enrichment, corrections, and infills of those missing in I-CeM. The BBCE already includes many additional codes to those in I-CeM, and can be linked to I-CeM through the individual identifiers (RecID) for each entrepreneur identified in the censuses to provide a complete resource of the census information on entrepreneurs for Britain as a whole. However, BBCE also has additional people from the census not recorded in I-CeM at all, or that cannot be identified in I-CeM. This paper gives the additional information on them.

This work was supported by the ESRC, and by a special ESRC financial supplementation for Scotland, for the project ES/M010953 ‘Drivers of Entrepreneurship and Small Businesses’. An overview of the project is provided in WP 1. WP 2 defines in detail the different censuses and the challenges they present for identifying entrepreneurs. WPs 3 and 4 describe how the data for 1851-81 and 1891-1911, respectively, were extracted from I-CeM and supplemented.

WP 20 summarises all these activities specifically for the Scottish data. A full list of Working Papers is given at the end of this paper. Full details of the BBCE database deposit at UKDA are given in the BBCE *User Guide*. The downloads with this paper give changes to specific coding and the new RecIDs which are additions to I-CeM.

2. Corrections and additions to I-CeM in BBCE

In the process of infilling data gaps in I-CeM various checks were run and from these a number of corrections were made based on inspection of the original scans of the Census Enumerators Books (CEBs). These corrections were identified either by the research team listed as authors here, or by research assistants working on the England and Wales, and Scottish records (see WP 3).¹ This Working Paper 21 describes the data codes for these corrections which allow near-replication of the analysis of BBCE in published research by the research team, as listed on the project's website:

<http://www.geog.cam.ac.uk/research/projects/driversofentrepreneurship>

The corrected codes make it possible for users to modify I-CeM to align with the data used in BBCE. However, the ESRC project used an improved version of I-CeM developed within Campop which is not identical with I-CeM version 1. The Campop data improvements will be part of I-CeM version 2, but additional improvements will also be made to that deposit (which will include many of the BBCE data corrections given here). As a result the data codes provided here allow near (but not perfect) replication of the main results of the BBCE data assembly and its core analysis using I-CeM version 1, reported in Bennett et al. (2019) *The Age of Entrepreneurship*, and elsewhere. Users seeking near-replication of the project's results should use the data codes given here; if there are any queries about precise relationship between the databases please contact the authors.

Most of the data corrections from the ESRC 'Drivers of Entrepreneurship and Small Businesses' are small in number, but affect most years, and often concern some of the most crucial people who otherwise become erroneous outliers or extreme cases; such as male

¹ Mark Latham, Gavin Robinson, Tiffany Shumaker and Rebecca Tyler for England and Wales; and Tobias Lunde, Annette Mackenzie and Amber Stevenson for Scotland.

entrepreneurs with large workforces who were sometimes mis-transcribed by FMP and hence in I-CeM as female.

In addition to coding corrections, a number of additional individuals in the census were identified who are not included in current versions of I-CeM. The main addition was for all entrepreneurs for 1871, which for England and Wales is a year not included in I-CeM at all but is included in BBCE for those who were entrepreneurs (See WP 12). As well as 1871, however, checks on the CEBs for missing or truncated transcriptions in I-CeM identified other individuals missing in I-CeM for 1851. The information and main data codes for these additional people are included in BBCE along with all the other entrepreneurs and their codes. Here additional data coding of these individuals is given so that other researchers will have the full information available that was used in the ESRC project and can thus replicate its results, and will be able to add these individuals separately to their analysis of I-CeM if desired.

3. New variables for corrected data coding

The new codes relate to two groups of variables:

- (1) Corrections of I-CeM for sex, age and marital status
- (2) For additional people I-CeM codes are given for number of servants, household size, and relationships within the household (I-CeM RELA codes). For RELA the simplified RELA10 code is given in BBCE, but in the downloads with this WP the full I-CeM RELA code is given.

Note that not every year will have every variable, e.g. there were no changes in Scotland for 1851. For 1871 all the new variables and codes are included directly in BBCE.

3.1. Corrections of I-CeM for sex, age and marital status

Variable: NEWSEX

Variable label: Gender

Type: Character

Max length: 1

Description: Corrected gender based on CEB checks.

Coverage: 1851-1861, 1881 England and Wales; 1861-1881 Scotland

Variable: NEWAGE

Variable label: Age

Type: Numeric

Max length: 4

Description: Corrected age based on CEB checks.

Coverage: 1851-1861, 1881 England and Wales

Variable: NEWMAR

Variable label: Marital Status

Type: Numeric

Max length: 4

Description: Corrected marital status based on CEB checks.

Coverage: 1861 England and Wales

3.2. Codes for additional people 1851 England and Wales not in I-CeM

Variable name and definitions as in I-CeM

Variable: SERVTS

Variable label: Total number of servants individual living with

Type: Numeric

Max length: 2

Description: Number of servants in the household.

Coverage: 1851 England and Wales only

Variable: HHS

Variable label: Size of household

Type: Numeric

Max length: 2

Description: Number of persons in the household; size of the household unit. The household consists of the head, their relatives and servants

Coverage: 1851 England and Wales only

Variable: RELA

Variable label: Relationship to household head code

Type: Numeric

Max length: 4

Description: Relationship to head of family, derived from relationship descriptor in CEBs, then coded using I-CeM dictionary (pp. 189-191).

Coverage: 1851 England and Wales only

4. A note on 1881 data

The ESRC project used two datasets for England and Wales 1881:

- (1) The pilot data derived from the Leverhulme pilot project. This used the 1881 GSU database deposited by Schürer and Woollard (2000). This predated I-CeM.
- (2) The I-CeM data.

The Leverhulme pilot project made checks on the CEBs resulting in some data corrections of sex, but no other demographic data. These data were used in some of the ESRC project analyses: e.g. of portfolios in farming (Radicic et al., 2017), and firm-size distribution (Monterbuno et al., 2019), and some subsequent research.

However, in the main ESRC project the core analysis used I-CeM data; e.g. as in Bennett et al. (2019a). The England and Wales 1881 downloads with this WP include corrections where the pilot demographic data differed from I-CeM. These are either based on the GSU data (which is considered a more accurate transcription) given in Schürer and Woollard (2000), or based on CEB checks during the pilot. As with other years, this only affects a small number of people, but includes one spurious code for a female employer with 400 employees, and one with 50 employees, which this download corrects to male.

Scotland in BBCE for 1881 was based solely on I-CeM data.

5. Conclusion

The variables presented in this working paper are the ones used by the BBCE project team. They consist of corrections to demographic data as included in I-CeM version 1, as well as a group of additional people missing from I-CeM for 1851 in England and Wales. It is expected that all or most of the corrections will be included in I-CeM version 2, but additional people will probably remain available only through the codings given in the uploads to this WP and in the BBCE.

Note that not every year has every variable, e.g. there are no changes for Scotland for 1851.

Researchers using these data will be able to achieve near-replication of the analysis reported in Bennett et al. (2019) *The Age of Entrepreneurship* and other publications.

Acknowledgments:

This research has been supported by the ESRC under project grant ES/M010953: **Drivers of Entrepreneurship and Small Businesses**. Piloting of the research for 1881 draws from Leverhulme Trust grant RG66385: **The long-term evolution of Small and Medium-Sized Enterprises (SMEs)** assisted by Gill Newton. An Isaac Newton Trust Grant, 18.40(g) **‘Business proprietor succession and firm size change 1851-1881’**, provided additional support to help track and code Scottish entrepreneurs. For 1871 acquisition of census data from S&N was supported by the main ESRC grant, with additional support for data coding and cleaning by Joe Day from Isaac Newton Trust research grant 17.07(d): **Business Employers in 1871**.

The census database used for 1851-61 and 1881-1911 derives from Higgs, Edward and Schürer, Kevin (University of Essex) (2014) *The Integrated Census Microdata (I-CeM) UKDA*, SN-7481; see also E. Higgs, C. Jones, K. Schürer and A. Wilkinson, *Integrated Census Microdata (I-CeM) Guide*, 2nd ed. (Colchester: Department of History, University of Essex, 2015). The census data for 1871 was provided by S&N; see WP 12.

We are very pleased to record our thanks for the help of research assistants who checked and filled out I-CeM truncated lines in 1861 using National Archive sources: Mark Latham,

Gavin Robinson, Tiffany Shumaker and Rebecca Tyler. Gavin Robinson also helped fill out 1851 I-CeM truncated lines using transcripts from S&N. In Scotland similar work with truncated lines at Scotland's People was assisted by Tobias Lunde, Annette Mackenzie and Amber Stevenson.

References

- Bennett, Robert J., Smith, Harry, van Lieshout, Carry, Montebruno, Piero and Newton, Gill (2019a) *The Age of Entrepreneurship: Business Proprietors, Self-employment and Corporations Since 1851*, Abingdon: Routledge.
- Bennett, R. J., Smith, H., Montebruno, P. (2019b) 'The Population of Non-corporate Business Proprietors in England and Wales 1891–1911', *Business History*, <https://doi.org/10.1080/00076791.2018.1534959>
- Higgs, Edward and Schürer, Kevin (2014) *The Integrated Census Microdata (I-CeM) UK Data Service*, SN-7481 [data collection]. <https://dx.doi.org/10.5255/UKDA-SN-7481-1>
- Higgs, E., Jones, C., Schürer, K. and Wilkinson, A. (2015), *Integrated Census Microdata (I-CeM) Guide*, Colchester: Department of History, University of Essex.
- Montebruno, P., Bennett, R.J., van Lieshout, C., Smith, H. (2019) *Research data supporting "A tale of two tails: Do Power Law and Lognormal models fit firm-size distributions in the mid-Victorian era?"*. Mendeley Data <http://doi:10.17632/86xkknmcw3.1>
- Radicić, D., Bennett, R.J. and Newton, G. (2017) Portfolio entrepreneurship in farming: Empirical evidence from the 1881 census for England and Wales. *Journal of Rural Studies*, 55, 289-302. DOI: [10.1016/j.jrurstud.2017.08.019](https://doi.org/10.1016/j.jrurstud.2017.08.019)
- Schürer, K., Higgs, E., Reid, A.M. and Garrett, E.M. (2016) *Integrated Census Microdata, 1851-1911, User Guide version V. 2 (I-CeM.2)*, Second edition, Colchester: Department of History, University of Essex.
- Schürer, Kevin and Woollard, Matthew (2000) *1881 Census for England and Wales, the Channel Islands and the Isle of Man (Enhanced Version)*, UKDA, University of Essex, SN-4177.

Other Working Papers:

Working paper series: ESRC project ES/M010953: '*Drivers of Entrepreneurship and Small Business*', University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure.

WP 1: Bennett, Robert J., Smith Harry J., van Lieshout, Carry, and Newton, Gill (2017) *Drivers of Entrepreneurship and Small Businesses: Project overview and database design*. <https://doi.org/10.17863/CAM.9508>

- WP 2: Bennett, Robert J., Smith Harry J. and van Lieshout, Carry (2017) *Employers and the self-employed in the censuses 1851-1911: The census as a source for identifying entrepreneurs, business numbers and size distribution*.
<https://doi.org/10.17863/CAM.9640>
- WP 3: van Lieshout, Carry, Bennett, Robert J., Smith, Harry J. and Newton, Gill (2017) *Identifying businesses and entrepreneurs in the Censuses 1851-1881*.
<https://doi.org/10.17863/CAM.9639>
- WP 4: Smith, Harry J., Bennett, Robert J., and van Lieshout, Carry (2017) *Extracting entrepreneurs from the Censuses, 1891-1911*. <https://doi.org/10.17863/CAM.9638>
- WP 5: Bennett, Robert J., Smith Harry J., van Lieshout, Carry, and Newton, Gill (2017) *Business sectors, occupations and aggregations of census data 1851-1911*.
<https://doi.org/10.17863/CAM.9874>
- WP 6: Smith, Harry J. and Bennett, Robert J. (2017) *Urban-Rural Classification using Census data, 1851-1911*. <https://doi.org/10.17863/CAM.15763>
- WP 7: Smith, Harry, Bennett, Robert J., and Radicic, Dragana (2017) *Classification of towns in 1891 using factor analysis*. <https://doi.org/10.17863/CAM.15767>
- WP 8: Bennett, Robert J., Smith, Harry, and Radicic, Dragana (2017) *Classification of occupations for economically active: Factor analysis of Registration Sub-Districts (RSDs) in 1891*. <https://doi.org/10.17863/CAM.15764>
- WP 9: Bennett, Robert, J., Montebruno, Piero, Smith, Harry, and van Lieshout, Carry (2018) *Reconstructing entrepreneurship and business numbers for censuses 1851-81*.
<https://doi.org/10.17863/CAM.37738>
- WP 9.2: Bennett, Robert, J., Montebruno, Piero, Smith, Harry, and van Lieshout, Carry (2019) *Reconstructing business proprietor responses for censuses 1851-81: a tailored logit cut-off method*.
- WP 10: Bennett, Robert, J., Smith, Harry and Radicic, Dragana (2018) *Classification of environments of entrepreneurship: Factor analysis of Registration Sub-Districts (RSDs) in 1891*. <https://doi.org/10.17863/CAM.26386>
- WP 11: Montebruno, Piero (2018) *Adjustment Weights 1891-1911: Weights to adjust entrepreneur numbers for non-response and misallocation bias in Censuses 1891-1911*. <https://doi.org/10.17863/CAM.26378>
- WP 12: van Lieshout, Carry, Day, Joseph, Montebruno, Piero and Bennett Robert J. (2018) *Extraction of data on Entrepreneurs from the 1871 Census to supplement I-CeM*.
<https://doi.org/10.17863/CAM.27488>
- WP 13: van Lieshout, Carry, Bennett, Robert J. and Smith Harry (2019) *Extracted data on employers and farmers compared with published tables in the Census General Reports, 1851-1881*. <https://doi.org/10.17863/CAM.37165>
- WP 14: van Lieshout, Carry, Bennett Robert J. and Montebruno, Piero (2019) *Company Directors: Directory and Census record linkage*.
<https://doi.org/10.17863/CAM.37166>
- WP 15: Bennett, Robert, J., Montebruno, Piero, Smith, Harry and van Lieshout, Carry (2019) *Entrepreneurial discrete choice: Modelling decisions between self-employment, employer and worker status*. <https://doi.org/10.17863/CAM.37312>
- WP 16: Satchell, M., Bennett, Robert J., Bogart, D. and Shaw-Taylor, L. (2019) *Constructing Parish-level Data and RSD-level Data on Transport Infrastructure in England and Wales 1851-1911*. <https://doi.org/10.17863/CAM.37313>
- WP 17: Satchell, M. and Bennett, Robert J. (2019) *Building a 1911 Historical Land Capacity GIS*. <https://doi.org/10.17863/CAM.42285>

- WP 18: Bennett, Robert, J., Smith, Harry, van Lieshout, Carry and Montebruno, Piero (2019) *Identification of business partnerships in the British population censuses 1851-1911 for BBCE*. <https://doi.org/10.17863/CAM.43890>
- WP 18: Bennett, Robert, J., Smith, Harry, van Lieshout, Carry and Montebruno, Piero (2019) *Identification of business partnerships in the British population censuses 1851-1911 for BBCE*. <https://doi.org/10.17863/CAM.43890>
- WP 19: Montebruno, Piero (2019) *Datasets and guide: downloads for reconstructing British census responses 1851-1881 for the BBCE*. <https://doi.org/10.17863/CAM.42285>
- WP 20: Smith, Harry, van Lieshout, Carry, Montebruno, Piero and Bennett, Robert, J. (2019) *Preparing Scottish census data in I-CeM for the British Business Census of Entrepreneurs (BBCE)*.
- WP 21: van Lieshout, Carry, Bennett, Robert, J., and Smith, Harry (2019) *Additional codes and people in the British Business Census of Entrepreneurs (BBCE) not available through I-CeM*.

Full list of all current Working Papers available at:

<http://www.geog.cam.ac.uk/research/projects/driversofentrepreneurship>